Markov Chains, Stochastic Processes, and Matrix Decompositions

Jack Gilbert

5 May 2014

Outline

1. Markov Chains
   - Probability Spaces, Random Variables and Expected Values
   - Markov Chains and Transition Matrices
   - The Chapman-Kologorov Equation
   - State Spaces
   - Recurrence and Irreducible Markov Chains

Outline

1. Markov Chains
   - Probability Spaces, Random Variables and Expected Values
   - Markov Chains and Transition Matrices
   - The Chapman-Kologorov Equation
   - State Spaces
   - Recurrence and Irreducible Markov Chains
2. Markov Chains and Linear Algebra
   - Introduction
   - Perron-Frobenius
   - Matrix Decompositions and Markov Chains
   - Spectral Representations

Outline

Probability Spaces

Definition

A probability space consists of three parts:

- A sample space Ω which is the set of all possible outcomes.
- A set of events F where each event is a set containing zero or more outcomes.
- A probability measure P which assigns events probabilities.

Definition

The sample space Ω of an experiment is the set of all possible outcomes of that experiment.

More Probability Spaces

Definition

An event is a subset of a sample space. An event A is said to occur if and only if the observed outcome $\omega \in A$.

Definition

If Ω is a sample space and if P is a function which associates a number for each event in Ω, then P is called the probability measure provided that:

- For any event $A, 0 \leq P(A) \leq 1$
- $P(\Omega) = 1$
- For any sequence $A_1, A_2, \ldots$ of disjoint events,

$$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i).$$

Random Variables

We will focus our attention on random variables, a key component of stochastic processes.

Definition

A random variable X with values in the set E is a function which assigns a value $X(\omega) \in E$ to each outcome $\omega \in \Omega$.

When E is finite, X is said to be a discrete random variable.

Definition

The discrete random variables $X_1, \ldots, X_n$ are said to be independent if
$P\{X_1 = a_1, \ldots, X_n = a_n\} = P\{X_1 = a_1\} \cdots P\{X_n = a_n\}$ for every $a_1, \ldots, a_n \in E$.

Markov Chains

With these basic definitions in hand, we can begin our exploration of Markov chains.

Definition

The stochastic process $X = \{X_n; n \in \mathbb{N}\}$ is called a Markov chain if $P\{X_{n+1} = j \mid X_0, \ldots, X_n\} = P\{X_{n+1} = j \mid X_n\}$ for every $j \in E, n \in \mathbb{N}$.

So a Markov chain is a sequence of random variables such that for any n, X_{n+1} is conditionally independent of $X_0, \ldots, X_{n-1}$ given X_n.

Properties of Markov Chains

Definition

The probabilities $P(i,j)$ are called the transition probabilities for the Markov chain X.

We can arrange the $P(i,j)$ into a square matrix, which will be critical to our understanding of stochastic matrices.

Definition

Let P be a square matrix with entries $P(i,j)$ where $i,j \in E$. P is called a transition matrix over E if

- For every $i,j \in E, P(i,j) \geq 0$
- For every $i \in E, \sum_{j \in E} P(i,j) = 1$.

Markov Notation

In general, the basic notation for Markov chains follows certain rules:

- $M(i,j)$ refers to the entry in row i, column j of matrix M.
- Column vectors are represented by lowercase letters, i.e. $f(i)$ refers to the i-th entry of column f.
- Row vectors are represented by Greek letters, i.e. $\pi(j)$ refers to the j-th entry of row π.

Example

The transition matrix for the set $E = \{1, 2, \dots\}$ is

$$P = \begin{bmatrix} P(0,0) & P(0,1) & P(0,2) & \cdots \\ P(1,0) & P(1,1) & P(1,2) & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

Transition Probabilities

Theorem

For every n, $m \in \mathbb{N}$ with $m \geq 1$ and $i_0, \ldots, i_m \in E$,

$$P\{X_{n+i} = i_1, \ldots, X_{n+m} = i_m \mid X_n = i_0\}$$
$$= P(i_0, i_1)P(i_1, i_2) \cdots P(i_{m-1}, i_m).$$

Corollary

Let π be a probability distribution on E. Suppose $P\{X_k = i_k\} = \pi(i_k)$ for every $i_k \in E$. Then for every $m \in \mathbb{N}$ and $i_0, \ldots, i_m \in E$,

$$P\{X_0 = i_0, X_1 = i_1, \ldots, X_m = i_m\}$$
$$= \pi(i_0)P(i_0, i_1) \cdots P(i_{m-1}, i_m).$$

The Chapman-Kolmogorov Equation

Lemma

For any $m \in \mathbb{N}$,

$$P\{X_{n+m} = j \mid X_n = i\} = P^m(i,j) \text{ for every } i,j \in E \text{ and } n \in \mathbb{N}.$$

In other words, the probability that the chain moves from state i to that j in m steps is the (i,j)th entry of the n-th power of the transition matrix P. Thus for any $m, n \in \mathbb{N}$,

$$P^{m+n} = P^m P^n$$

which in turn becomes

$$P^{m+n}(i,j) = \sum_{k \in E} P^m(i,k) P^n(k,j);\, i,j \in E.$$

This is called the Chapman-Kolmogorov equation.

The Chapman-Kolmogorov Equation

Example

Let $X = \{X_n; n \in \mathbb{N}\}$ be a Markov chain with state space $E = \{a, b, c\}$ and transition matrix

$$P = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{2}{3} & 0 & \frac{1}{3} \\ \frac{3}{5} & \frac{2}{5} & 0 \end{bmatrix}$$

Then

$$P\{X_1 = b, X_2 = c, X_3 = a, X_4 = c, X_5 = a, X_6 = c, X_7 = b \mid X_0 = c\}$$
$$= P(c, b)P(b, c)P(c, a)P(a, c)P(c, a)P(a, c)P(c, b)$$

The Chapman-Kolmogorov Equation

Example (Continued)

$$= \frac{2}{5} \cdot \frac{1}{3} \cdot \frac{3}{5} \cdot \frac{1}{4} \cdot \frac{3}{5} \cdot \frac{1}{4} \cdot \frac{2}{5}$$
$$= \frac{3}{2500}.$$

The two-step transition probabilities are given by

$$P^2 = \begin{bmatrix} \frac{17}{30} & \frac{9}{40} & \frac{5}{24} \\ \frac{8}{15} & \frac{3}{10} & \frac{1}{6} \\ \frac{17}{30} & \frac{3}{20} & \frac{17}{60} \end{bmatrix}$$

where in this case $P\{X_{n+2} = c \mid X_n = b\} = P^2(b, c) = \frac{1}{6}$.

State Spaces

Definition

Given a Markov chain X, a state space E, and a transition matrix P, let T be the time of the first visit to state j and let N_j be the total visits to state j. Then

- State j is recurrent if $P_j\{T < \infty\} = 1$. Otherwise, j is transient if $P_j\{T = +\infty\} > 0$.
- A recurrent state j is null if $E_j\{T\} = \infty$; otherwise j is non-null.
- A recurrent state j is periodic with period δ if $\delta \geq 2$ is the largest integer for $P_j\{T = n\delta \text{ for some } n \geq 1\} = 1$.
- A set of states is closed if no state outside the set can be reached from within the set.

State Spaces

Definition (Continued)

- A state forming a closed set by itself is called an absorbing state.
- A closed set is irreducible if no proper subset of it is closed.
- Thus a Markov chain is irreducible if its only closed set is the set of all states.

State Spaces

Example

Consider the Markov chain with state space $E = \{a, b, c, d, e\}$ and transition matrix

$$P = \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{4} & 0 & \frac{3}{4} & 0 \\ 0 & 0 & \frac{1}{3} & 0 & \frac{2}{3} \\ \frac{1}{4} & \frac{1}{2} & 0 & \frac{1}{4} & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{3} \end{bmatrix}$$

The closed sets are $\{a, b, c, d, e\}$ and $\{a, c, e\}$. Since there exist two closed sets, the chain is not irreducible.

State Spaces

Example (Continued)

By deleting the second and fourth rows and column we end up with the matrix

$$Q = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{3} & \frac{2}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}$$

which is the Markov matrix corresponding to the restriction of X to the closed set $\{a, c, e\}$. We can rearrange P for easier analysis as such:

$$P^* = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & \frac{1}{3} & \frac{2}{3} & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{4} & \frac{3}{4} \\ \frac{1}{4} & 0 & 0 & \frac{1}{2} & \frac{1}{4} \end{bmatrix}$$

Recurrence and Irreducibility

Theorem

From a recurrent state, only recurrent states can be reached.

So no recurrent state can reach a transient state and the set of all recurrent states is closed.

Lemma

For each recurrent state j there exists an irreducible closed set C which includes j.

Proof.

Let j be a recurrent state and let C be the set of all states which can be reached from j. Then C is a closed set. If $i \in C$ then $j \to i$. Since j is recurrent, our previous lemma implies that $i \to j$. There must be some state k such that $j \to k$, and thus $i \to k$. $\square$

Recurrence and Irreducibility

Theorem

In a Markov chain, the recurrent states can be divided uniquely into irreducible closed sets $C_1, C_2, \ldots$

Using this theorem we can arrange our transition matrix in the following form

$$P = \begin{bmatrix} P_1 & 0 & 0 & \cdots & 0 \\ 0 & P_2 & 0 & \cdots & 0 \\ 0 & 0 & P_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ Q_1 & Q_2 & Q_3 & \cdots & Q \end{bmatrix}$$

where $P_1, P_2, \ldots$ are the Markov matrices corresponding to sets $C_1, C_2, \ldots$ of states and each C_i is an irreducible Markov chain.

Recurrence and Irreducibility

The following tie these ideas together.

Theorem

Let X be an irreducible Markov chain. Then either all states are transient, or all are recurrent null, or all are recurrent non-null. Either all states are aperiodic, or else all are periodic with the same period δ.

Corollary

Let C be an irreducible closed set with finitely many states. Then no state in C is recurrent null.

Corollary

If C is an irreducible closed set with finitely many states, then C has no transient states.

Markov Chains and Linear Algebra

This theorem gives some of the flavor of working with Markov chains in a linear algebra setting.

Theorem

Let X be an irreducible Markov chain. Consider the system of linear equations

$$\vec{v}(j) = \sum_{i \in E} \vec{v}(i) P(i,j), j \in E.$$

Then all states are recurrent non-null if and only if there exists a solution $\vec{v}$ with $\sum_{j \in E} \vec{v}(j) = 1$.

If there is a solution $\vec{v}$ then $\vec{v}(j) > 0$ for every $j \in E$, and $\vec{v}$ is unique.

The Perron-Frobenius Theorems

Theorem (Perron-Frobenius Part 1)

Let A be a square matrix of size n with non-negative entries. Then

- *A has a positive eigenvalue λ_0 with left eigenvector $\vec{x_0}$ such that $\vec{x_0}$ is non-negative and non-zero.*
- *If λ is any other eigenvalue of $A, \mid \lambda \mid \leq \lambda_0$.*
- *If λ is an eigenvalue of A and $\mid \lambda \mid = \lambda_0$, then $\mu = \frac{\lambda}{\lambda_0}$ is a root of unity and $\mu^k \lambda_0$ is an eigenvalue of A for $k = 0, 1, 2, \ldots$*

The Perron-Frobenius Theorems

Theorem (Perron-Frobenius Part 2)

Let A be a square matrix of size n with non-negative entries such that A^m has all positive entries for some m. Then

- *A has a positive eigenvalue λ_0 with a corresponding left eigenvector $\vec{x_0}$ where the entries of $\vec{x_0}$ are positive.*
- *If λ is any other eigenvalue of $A, \mid \lambda \mid < \lambda_0$.*
- *λ_0 has multiplicity 1.*

The Perron-Frobenius Theorems

Corollary

Let P be an irreducible Markov matrix. Then 1 is a simple eigenvalue of P. For any other eigenvalue λ of P we have $\mid \lambda \mid \leq 1$. If P is aperiodic then $\mid \lambda \mid < 1$ for all other eigenvalues of P. If P is periodic with period δ then there are δ eigenvalues with an absolute value equal to 1. These are all distinct and are

$$\lambda_1 = 1, \lambda_2 = c, \ldots, \lambda_\delta = c^{\delta-1}; c = e^{2\pi i/\delta}.$$

A Perron-Frobenius Example

Example

$$P = \begin{bmatrix} 0 & 0 & 0.2 & 0.3 & 0.5 \\ 0 & 0 & 0.5 & 0.5 & 0 \\ 0.4 & 0.6 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0.2 & 0.8 & 0 & 0 & 0 \end{bmatrix}$$

P is irreducible and periodic with period $\delta = 2$. Then

$$P^2 = \begin{bmatrix} 0.48 & 0.52 & 0 & 0 & 0 \\ 0.70 & 0.30 & 0 & 0 & 0 \\ 0 & 0 & 0.38 & 0.42 & 0.20 \\ 0 & 0 & 0.20 & 0.30 & 0.50 \\ 0 & 0 & 0.44 & 0.46 & 0.10 \end{bmatrix}$$

A Perron-Frobenius Example

Example (Continued)

The eigenvalues of the 2 by 2 matrix in the top-left corner of P^2 are 1 and -0.22. Since $1, -0.22$ are eigenvalues of P, their square roots will be eigenvalues for $P : 1, -1, i\sqrt{0.22}, -i\sqrt{0.22}$. The final eigenvalue must go into itself by a rotation of 180 degrees and thus must be 0.

An Implicit Theorem

Theorem

All finite stochastic matrices P have 1 as an eigenvalue and there exist non-negative eigenvectors corresponding to $\lambda = 1$.

Proof.

Since each row of P sums to 1, $\vec{y}$ is a right eigenvector. Since all finite chains have at least one positive persistent state, we know there exists a closed irreducible subset S and the Markov chain associated with S is irreducible positive persistent. We know for S there exists an invariant probability vector. Assume P is a square matrix of size n and rewrite P in block form with

$$P^* = \begin{bmatrix} P_1 & 0 \\ R & Q \end{bmatrix}$$

An Implicit Theorem

Proof (Continued).

P_1 is the probability transition matrix corresponding to S. Let $\vec{\pi} = (\pi_1, \pi_2, \ldots, \pi_k)$ be the invariant probability vector for P_1. Define $\vec{\gamma} = (\pi_1, \pi_2, \ldots, \pi_k, 0, 0, \ldots, 0)$ and note $\vec{\gamma}P = \vec{\gamma}$. Hence $\vec{\gamma}$ is a left eigenvector for P corresponding to $\lambda = 1$. Additionally,

$$\sum_{i=1}^{n} \gamma_i = 1.$$

$\square$

This shows that $\lambda = 1$ is the largest possible eigenvalue for a finite stochastic matrix P.

Another Theorem

Theorem

If P is the transition matrix for a finite Markov chain, then the multiplicity of the eigenvalue 1 is equal to the number of irreducible subsets of the chain.

Proof (First Half).

Arrange P based on the irreducible subsets of the chain $C_1, C_2, \ldots$.

$$P = \begin{bmatrix} P_1 & 0 & \cdots & 0 & 0 \\ 0 & P_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & P_m & \vdots \\ R_1 & R_2 & \cdots & R_m & Q \end{bmatrix}$$

Another Theorem

Proof (First Half Continued).

Each P_k corresponds to a subset C_k for an irreducible positive persistent chain. The $\vec{x_i}$s for each C_i are linearly independent; thus the multiplicity for $\lambda = 1$ is at least equal to the number of subsets C. □

An Infinite Decomposition

Theorem

Let $\{X_t; t = 0, 1, 2, \dots\}$ be a Markov chain with state $\{0, 1, 2, \dots\}$ and a transition matrix P. Let P be irreducible and consist of persistent positive recurrent states. Then $I - P = (A - I)(B - S)$ where

- *A is strictly upper triangular with $a_{ij} = E_i\{\text{number of times } X = j \text{ before } X \text{ reaches } \Delta_{j-i}\}, i < j$.*
- *B is strictly lower triangular with $b_{ij} = P_i\{X^i = j\}, i > j$.*
- *S is diagonal where $s_j = \sum_{j=0}^{i-1} b_{ij}$ (and $s_0 = 0$). Moreover, $a_{ij} < \infty$ and $i < j$.*

Spectral Representations

Suppose we have a diagonalizable matrix A. Define B_k to be the matrix obtained by multiplying the column vector f_k with the row vector π_k where f_k, π_k are from A. Explicitly,

$$B_k = f_k \pi_k.$$

Then we can represent A in the following manner:

$$A = \lambda_1 B_1 + \lambda_2 B_2 + \cdots + \lambda_n B_n.$$

This is the spectral representation of A, and it holds some key properties relevant to our discussion of Markov chains. For example, for the k-th power of A we have

$$A^k = \lambda_1^k B_1 + \cdots + \lambda_n^k B_n.$$

A Final Example

Example

Let

$$P = \begin{bmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{bmatrix}$$

$\lambda_1 = 1$; $\lambda_2 = .5$. We can now calculate B_1:

$$B_1 = f_1 \pi_1 = \begin{bmatrix} 0.6 & 0.4 \\ 0.6 & 0.4 \end{bmatrix}$$

Since $P^0 = I = B_1 + B_2$ for $k = 0$,

$$B_2 = \begin{bmatrix} 0.4 & -0.4 \\ -0.6 & 0.6 \end{bmatrix}$$

A Final Example

Example (Continued)

Thus the spectral representation for P^k is

$$P^k = \begin{bmatrix} 0.6 & 0.4 \\ 0.6 & 0.4 \end{bmatrix} + (0.5)^k \begin{bmatrix} 0.4 & -0.4 \\ -0.6 & 0.6 \end{bmatrix}, k = 0, 1, \ldots$$

The limit as $k \to \infty$ has $(0.5)^k$ approach zero, so

$$P^\infty = \lim_k P^k = \begin{bmatrix} 0.6 & 0.4 \\ 0.6 & 0.4 \end{bmatrix}.$$

Sources

R. Beezer.
A Second Course In Linear Algebra.
Open-License, 2014.

E. Cinlar.
Introduction to Stochastic Processes.
Pretence-Hall, 1975.

D. Heyman.
A Decomposition Theorem for Infinite Stochastic Matrices.
J. Appl. Prob. 32, 1995.

D. Isaacson.
Markov Chains: Theory and Applications.
John Wiley and Sons, 1976.